JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 1998, p. 486–492 Vol. 36, No. 2
0095-1137/98/$04.00+0


# Molecular Fingerprinting of *Mycobacterium tuberculosis* and Risk Factors for Tuberculosis Transmission in Paris, France, and Surrounding Area

M. C. GUTIÉRREZ,[1] V. VINCENT,[1]* D. AUBERT,[2] J. BIZET,[3] O. GAILLOT,[4] L. LEBRUN,[5] C. LE PENDEVEN,[6] M. P. LE PENNEC,[7] D. MATHIEU,[8] C. OFFREDO,[9] B. PANGON,[10] AND C. PIERRE-AUDIGIER[11]

*Centre National de Référence des Mycobactéries, Institut Pasteur, 75724 Paris,[1] Service de Microbiologie, Hôpital Max Fourestier, 92000 Nanterre,[2] Laboratoire de Microbiologie, Hôpital des Prisons de Fresnes, 94261 Fresnes,[3] Service de Microbiologie, Hôpital Antoine Béclère, 92141 Clamart,[5] Laboratoire Départemental de Seine Saint Denis, 93140 Bondy,[6] Service de Microbiologie, Hôpital Robet Ballanger, 93602 Aulnay sur Bois,[7] Service de Microbiologie, Hôpital Paul Brousse, 94800 Villejuif,[8] Service de Microbiologie, Hôpital Laënnec, 75007 Paris,[9] Centre Hospitalier de Versailles, 78000 Versailles,[10] Service de Microbiologie, Hôpital Necker-Enfants-Malades, 75015 Paris,[11] and Service de Microbiologie, Hôpital Boucicaut, 75015 Paris,[4] France*

Received 7 August 1997/Returned for modification 1 October 1997/Accepted 4 November 1997

**Forty-three percent of the tuberculosis cases reported in France are from the Ile de France region. The incidence of tuberculosis in this region is 33 cases per 100,000 inhabitants, twice the national average. A restriction fragment length polymorphism (RFLP) analysis was performed with clinical isolates of *Mycobacterium tuberculosis* isolated during 1995 in 10 hospitals in Paris and surrounding areas to detect tuberculosis transmission and define the factors associated with clustering in this population. The molecular markers used were the insertion sequence IS*6110* and the direct repeat (DR) sequence. Social, demographic, and clinical data were collected from the patients' medical files. Ten patients with isolates with a single copy of IS*6110* were excluded from further analysis. Twenty-four patients with false-positive cultures due to laboratory contamination (based on RFLP analysis with IS*6110* and examination of patient data) were also excluded. The study was then conducted with 272 strains isolated from 272 patients. Further fingerprinting was performed by using the DR element with strains with patterns by RFLP analysis with IS*6110* that differed by one band only and strains with identical patterns by RFLP analysis with IS*6110* and with low numbers of copies of IS*6110*. The combined use of both markers identified unique patterns for 177 strains and clustered 95 (35.7%) strains in 26 groups, each containing isolates from 2 to 12 patients. The clustering was strongly associated with homelessness and the male sex. It was not associated with age, birth in a foreign country, human immunodeficiency virus positivity, or residence in hostels or prison. Isolates from homeless people were often included in large clusters, and homeless people could be the source of tuberculosis transmission for more than 50% of the clustered patients. These results suggest that homeless people play a key role in the spread of *M. tuberculosis* in the community and that poor socioeconomic conditions are the main risk factors associated with active tuberculosis transmission.**

Since the late 1980s, tuberculosis has resurged in developed countries. It is important to consider whether the disease results from reactivation of a previous infection or from active transmission and to understand the risk factors associated with the transmission of tuberculosis because it makes possible the optimal use of health care resources and the selection of top priorities for tuberculosis control (1, 28).

It was estimated by conventional epidemiologic methods that 90% of the active cases of tuberculosis in developed countries resulted from reactivation during adulthood of an infection contracted years before and that recently transmitted disease had a minor role (2). Classical bacteriologic methods are of limited practical value because they distinguish relatively few *Mycobacterium tuberculosis* strains (16). In the last few years, various repetitive genetic elements that contribute to the polymorphism of *M. tuberculosis* DNA have been described. These include the insertion sequence IS*6110* (31), the direct repeat (DR) sequence (17), the polymorphic GC-rich repetitive sequences (25), the major polymorphic tandem repeats (18), and the $(GTG)_5$ oligonucleotides (38). Recent studies have demonstrated the value of these elements for epidemiologic studies for tuberculosis control. Analysis of the distribution of the conserved insertion sequence IS*6110* by restriction fragment length polymorphism (RFLP) analysis is the most common method for distinguishing *M. tuberculosis* strains (9, 22, 26, 33, 35, 37, 40, 42, 43).

Studies that use this sequence to trace outbreaks in the hospital or in the community have shown that epidemiologically linked strains of *M. tuberculosis* have identical patterns by RFLP analysis with IS*6110*, whereas unrelated strains have different patterns (7, 10). However, RFLP analysis with IS*6110* alone may be inconclusive for strains carrying few copies of IS*6110* (17, 43) and for strains presenting similar but not identical patterns since this insertion sequence may be mobile (10, 11, 14, 35). Use of a secondary genetic marker is necessary to discriminate among such isolates (8).

Recent epidemiologic studies that combine population-

* Corresponding author. Mailing address: Centre National de Référence des Mycobactéries, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris cedex 15, France. Phone: (33) (1) 45688360. Fax: (33) (1) 40613118. E-mail: vvincent@pasteur.fr.

based methods and molecular biology-based methods (1, 3, 15, 28, 41) have suggested that patients in developed countries whose *M. tuberculosis* strains have identical RFLP patterns could be classified into epidemiologically linked clusters, although few cases of transmission have been identified by conventional contact tracing. Clustering reflects recent transmission or a common source of infection. These combined studies have shown that recent transmission contributes substantially to the increase in the incidence of tuberculosis in some areas (3, 41). The investigation of the microbiological, clinical, social, and demographic factors associated with these epidemiologically related cases is of great importance for defining those risk factors contributing to patients' being part of the chain of transmission of tuberculosis (1, 28).

The incidence of tuberculosis in France has increased in the same way as it has in other developed countries. After a long period of a decreased incidence of tuberculosis the national rate remained fairly constant from 1986 to 1991. Then it increased from 14.7 cases per 100,000 people in 1991 to 16.6 cases per 100,000 people in 1994. The increase mostly affected urban areas, especially those in the Ile de France region. This region encompasses eight departments; one department corresponds to the city of Paris, France, and the others correspond to the area surrounding Paris. The region, which contains only 19% of the French population, has 42.8% of all cases of tuberculosis reported in France. The regional incidence of tuberculosis is more than twice the national average. Two departments from this region (Paris and Seine Saint Denis) have the highest incidences, more than 3.5 times the national average in Paris and 3 times the national average in Seine Saint Denis (6, 30).

The aim of this prospective study was to detect active tuberculosis transmission and to identify the risk factors associated with the transmission of tuberculosis in patients from the Ile de France region over a 1-year period (1995). We used systematic IS*6110*- and DR sequence-based DNA fingerprinting and univariate and multivariate analyses of social, demographic, and clinical information about the patients enrolled in the study.

## MATERIALS AND METHODS

**Patient population and clinical bacterial isolates.** The study population comprised all patients from whom at least one sample positive for *M. tuberculosis* by culture was collected between 1 January 1995 and 31 December 1995 in the microbiological laboratories of 10 medical institutions located in Paris or the surrounding area. In addition to those located in Seine Saint Denis, these medical institutions were hospitals with specialized pulmonary units, a pediatric hospital, a prison hospital, and hospitals dealing with large number of homeless or human immunodeficiency virus (HIV)-positive patients.

All isolates were identified as *M. tuberculosis* by the participating laboratories. For each patient included in the survey, one strain of *M. tuberculosis* and an anonymously coded form containing patient data were sent to the National Reference Center for Mycobacteria at the Institut Pasteur. The data were collected from the medical file records by bacteriologists. They included age, sex, country of birth, type of residence, HIV serological status, and pathology associated with tuberculosis. The specific information collected about tuberculosis included the date of disease onset, history of previous tuberculosis, site(s) of the disease, and the following microbiological data: date of sampling, date of specimen processing, smear microscopy results, number of colonies grown on solid medium, and strain identification.

**DNA probes.** Two different DNA probes were used. The IS*6110* probe is an 868-bp, PCR-amplified probe containing the sequence of the right arm of the IS*6110* (32) from the *M. tuberculosis* Mt 14323 reference strain (34). The DR sequence probe is a 36-bp oligonucleotide used to detect the direct repeats in the hot-spot integration region of IS*6110* (17). The probes were labeled nonradioactively and were detected with an enhanced chemiluminescence kit (ECL Amersham International).

**Southern blot analysis.** Mycobacterial strains were cultured and chromosomal DNA was extracted as described previously (32). IS*6110* RFLP analysis was performed by a standard method (34). RFLP analysis with the DR sequence was performed by reprobing the membranes with the DR sequence as described previously (17).

**Fingerprint analysis.** The autoradiographs of the blots probed with IS*6110* were scanned with a ScanJet II cx (Hewlett-Packard) scanner, and the patterns were analyzed by using the Taxotron package (Taxolab Software; Institut Pasteur) comprising the RestrictoScan, RestrictoTyper, Adanson, and Dendrograph programs. The bands were detected with the RestrictoScan program. Fragment lengths were calculated by using the Schaffer-Sederoff algorithm. The degree of similarity of the fingerprint patterns was calculated by using the RestrictoTyper program, with a linear error tolerance of between 3.5 and 5%, proportional to the size of the bands. The relationships between isolates were assessed by the unweighted pair group method of averages (19) by the Adanson program, and a dendrogram was generated by the Dendrograph program. The patterns that were identified as similar by computer analysis were compared visually. The strains were classified as clustered if the numbers and molecular sizes of the bands of the IS*6110*-RFLP patterns were identical. Strains with a single copy of IS*6110* were excluded from the study and were not analyzed further.

The patterns obtained by fingerprinting with the DR sequence were compared by visual examination. They were performed (i) for strains with six or more IS*6110* elements which differed by one band only (an additional band or a band with a different molecular size), (ii) for strains with identical IS*6110*-RFLP patterns composed of two to five bands, and (iii) for most of the strains included in clusters with six or more patients to check that the identity of the patterns was not due to the typing limits of RFLP analysis using IS*6110*.

Strains were considered to belong to a cluster if they had (i) indistinguishable IS*6110*-RFLP patterns patterns with six or more fragments, (ii) IS*6110*-RFLP patterns with six or more bands that differed by only one copy and indistinguishable DR-RFLP patterns, or (iii) indistinguishable IS*6110*-RFLP patterns with two to five bands and identical DR-RFLP patterns. Isolates with unique fingerprints were considered unclustered.

**Statistical analysis.** Data were entered by using EpiSurv 1.99 Fr.v (Epiconcept, Paris, France) and were analyzed with EpiInfo 5.01 Fr.v (Centers for Disease Control and Prevention and World Health Organization) programs with the support of Epicentre, Paris, France.

Patients born in French overseas departments were included in the group of foreign-born patients. Age was categorized as either a dichotomous variable (less than 60 years or 60 years and older) or as a continuous variable grouped into 10-year categories. The type of residence was coded into four categories: individual or family, hostel for immigrant workers, jail, and homeless. The site of tuberculous disease was classified as pulmonary (pulmonary or pulmonary plus extrapulmonary) or exclusively extrapulmonary.

Patients were divided into two groups. Group 1 contained all the patients with unclustered isolates, and group 2 contained all patients with clustered isolates. Chi-square tests were performed to test the univariate risk factors for belonging to a cluster. When cell sizes were expected to be smaller than five, Fisher's exact test was used. Adjusted odds ratios with 95% confidence intervals were calculated.

Variables identified by univariate analysis to be significantly associated with clustering along with those identified in other studies (1, 28, 40) were included in a multivariate logistic regression model, with "clustered" and "unclustered" being the dependent variables.

## RESULTS

**Strains included in the RFLP analysis.** During 1995, a total of 327 isolates obtained in the 10 microbiological laboratories participating in this study were sent to Institut Pasteur. Two strains from five patients were received independently from two laboratories. Only one strain was processed for each patient. Sixteen isolates were nonviable or contaminated and could not be used for RFLP analysis. The RFLP patterns of 306 isolates from 306 patients were determined.

**Detection of laboratory cross contamination.** Microbiological laboratories sent to the Institut Pasteur 12 strains noted as presumed laboratory contaminations. These strains were from unexpectedly positive cultures from Ziehl-Neelsen-negative specimens collected from patients without other cultures positive for *M. tuberculosis* and whose clinical course was not likely to have resulted from tuberculosis. The strains were considered to be isolated as a result of a cross contamination if their RFLP patterns were the same as that of a true-positive specimen processed concurrently. RFLP analysis confirmed the suspicion of cross contamination for 11 strains from seven independent series of specimen processing procedures. The remaining strain had a unique pattern by RFLP analysis with IS*6110*, and thus laboratory contamination could be excluded.

Moreover, we considered the possibility of laboratory cross contamination for strains isolated from microscopically nega-

TABLE 1. Description of patients with strains with unique RFLP patterns and 59 patients in 10 clusters of four or more patients (clusters A to J)

| Group | No. of patients | No. of IS*6110* copies | Median age (yr) | Percent | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Male | Foreign-born | HIV positive | Residence | | | | Alcohol use |
| | | | | | | | Homeless | Hostel[a] | Prison | Private | |
| Unique RFLP pattern[b] | 177 | | 43 | 62 | 60 | 17 | 10 | 9 | 13 | 68 | 9 |
| Cluster | | | | | | | | | | | |
| A | 12 | 8–9 | 47 | 83[c] | 33 | 0 | 75[e] | 8 | 0 | 17 | 36[c] |
| B | 8 | 13 | 49 | 100[c] | 86 | 0 | 50[d] | 0 | 0 | 37 | 25 |
| C | 7 | 8 | 47 | 100[c] | 29 | 0 | 57[c] | 0 | 0 | 29 | 57[d] |
| D | 6 | 12 | 50 | 83 | 33 | 0 | 50[c] | 17 | 0 | 33 | 33 |
| E | 5 | 14–15 | 31 | 80 | 100 | 25 | 20 | 80[e] | 0 | 0 | 0 |
| F | 5 | 10 | 37 | 80 | 80 | 0 | 0 | 0 | 20 | 40 | 20 |
| G | 4 | 9 | 35 | 50 | 25 | 25 | 0 | 0 | 0 | 75 | 0 |
| H | 4 | 10 | 39 | 75 | 75 | 0 | 0 | 25 | 25 | 25 | 0 |
| I | 4 | 8 | 50 | 50 | 25 | 0 | 0 | 0 | 50 | 25 | 0 |
| J | 4 | 6–7 | 52 | 75 | 25 | 25 | 0 | 0 | 0 | 100 | 0 |

[a] Hostel for immigrant workers.
[b] Means or percentages for patients with strains with unique RFLP patterns were calculated by excluding the data for the patients for whom specific data were not available.
[c] $P < 0.05$ ($P$ values are determined by comparison with characteristics of patients with strains with unique RFLP patterns).
[d] $P < 0.01$.
[e] $P < 0.001$.

tive specimens included in clusters containing strains from smear-positive specimens processed at the same laboratory. The dates of processing for all specimens from the true-positive clustered patients were checked by the participating laboratories. Smear-negative specimens processed concurrently were classified as contaminated. These specimens were from 13 patients, mainly HIV-positive patients, neoplasic patients, or patients with previous tuberculosis for whom cross contamination had not been suspected.

In total, 24 specimens were considered to be contaminated with 12 different strains of *M. tuberculosis* from five laboratories. The contamination rate, i.e., the number of patients with false-positive cultures versus the number of patients with genuine *M. tuberculosis* cultures, was 7.8%. Both the strains and the patients involved in laboratory cross contamination were excluded from further analysis.

**DNA polymorphisms of *M. tuberculosis* strains.** A total of 282 IS*6110*-based profiles for strains from 282 patients were studied after the exclusion of false-positive cultures. Of the 224 different patterns found, 195 were observed only once and 29 were shared by two or more isolates. The rate of diversity of the patterns obtained by RFLP analysis with IS*6110* for the *M. tuberculosis* strains included in this survey was 79.4%.

The number of copies of IS*6110* per isolate ranged from 1 to 20 (mean, 10). Most isolates (92.5%) carried six or more IS*6110* copies. Accordingly, 10 isolates with single-band patterns were excluded from subsequent analysis, along with the 10 corresponding patients. Previous studies concluded that strains of *M. tuberculosis* with only one copy of IS*6110* may correspond to epidemiologically unrelated cases of tuberculosis (17, 43). We considered that the DR-based fingerprint from *Pvu*II-digested DNAs is a good secondary marker which, contrary to *Alu*I patterns, may not be discriminatory enough to be used as a unique molecular marker (36).

The *M. tuberculosis* isolates from 272 patients were classified as either independent or clustered. The DR-RFLP pattern was used as a secondary molecular marker to support or rule out strain clustering, as detailed in Materials and Methods. The clustering of 20 strains with unique IS*6110* patterns with six or more copies that differed from another pattern by only one band was confirmed on the basis of identical DR-RFLP patterns. Distinct clusters of isolates by RFLP analysis with IS*6110* with patterns differing by only one band were merged into a single cluster according to their identical profiles by RFLP analysis with the DR sequence. Clusters A and E resulted from such fusions (Table 1). Two strains with an identical three-band IS*6110*-RFLP pattern and different DR-RFLP profiles were classified as unclustered. For clusters with six or more patients, the DR-based patterns of all 21 strains analyzed were identical within each IS*6110*-based cluster.

The combination of the patterns obtained with the two markers gave a rate of 74.26% strain diversity. A total of 177 patients (65.1%) had isolates with unique fingerprints and were considered to have independent cases of infection. The other 95 patients (34.9%) had strains belonging to 1 of the 26 clusters, suggesting that they belonged to groups of individuals among whom tuberculosis was transmitted. Clusters contained from 2 to 12 patients each (mean, 3.6 patients). Fourteen clusters contained more than two patients each (Fig. 1).

There were statistically significant differences in the IS*6110* copy numbers carried by strains isolated from foreign-born and native patients. Strains with a high IS*6110* copy number (more than 15) were isolated more frequently from patients born outside metropolitan France ($P = 0.04$; 6.9% of foreign-born versus 1.1% of French patients; Fig. 2). Strains carrying more than 15 copies of IS*6110* were significantly less clustered than strains carrying 15 or fewer copies of IS*6110* ($P = 0.007$; 0% for more than 15 copies and 40.1% for 15 or fewer copies). No other statistically significant differences were found between foreign-born and French patients.

**Characteristics of the patients.** The information available for the 272 patients showed that 40.9% were born in France, 12.3% were born in French overseas departments, and 56.8% were born in 31 foreign countries. One-third of the foreign-born patients were from sub-Saharan African countries, one-third were from northern Africa, 24 were from European countries other than France, 11 were from Asia, and 1 was from Brazil.

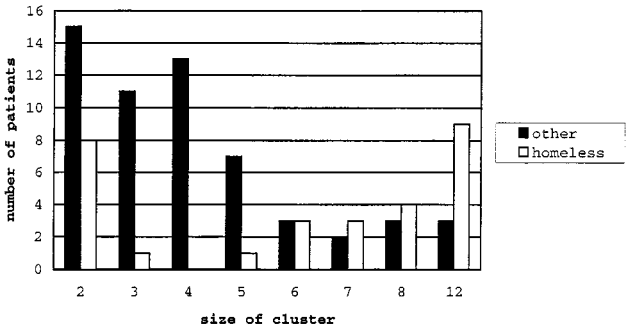


FIG. 1. Distribution of size of clusters among homeless and nonhomeless clustered patients. Data for patients whose residence is unknown are not included.

The patients were from 8 to 93 years old, with a mean age of 42.5 years and a standard deviation of 17.7 years. Most of the patients (80%) were between 20 and 65 years of age, and 70% of the patients were male.

Residence was classified as private homes for 136 patients, while 43 patients were homeless, 29 patients were in prison, and 20 patients were living in hostels for immigrant workers.

Of the patients whose HIV status was recorded, 31 were seropositive and 162 were seronegative at the time of tuberculosis diagnosis. Alcohol use was recorded for 34 patients.

Most patients (73%) presented with strictly pulmonary tuberculosis, and 14.5% presented with strictly extrapulmonary disease, mainly tuberculous lymphadenitis. The remaining patients presented with combined pulmonary and extrapulmonary tuberculosis. A total of 41.7% of the 230 patients with pulmonary tuberculosis had positive microscopy smears. Six patients presenting with tuberculous lymphadenitis belonged to clusters; three of these patients were HIV seropositive.

A previous history of tuberculosis was documented for 30 patients. Among these patients, 17 were included in 11 clusters. In four instances, several patients with a previous diagnosis of tuberculosis belonged to single clusters, indicating that reactivation was most unlikely.

**Cluster analysis.** Table 1 presents the characteristics of the patients and strains in 10 clusters each containing four or more patients. Nosocomial transmission was considered for one patient in cluster A who was living at home. This patient had been hospitalized several times before the diagnosis of tuberculosis in the same hospital and at the same time as two of the homeless patients in cluster A. Most of the patients in cluster B were born in Maghreb countries. Most of the patients in cluster C were alcoholic and/or homeless individuals or lived in the same department where the homeless patients were found. Cluster E included five patients, all of whom were born in Senegal or Mali, and these patients were younger (23 to 37 years old) than the average for the study population. Four men among these patients lived in hostels for immigrant workers in the Seine Saint Denis department, and one woman was homeless. Three patients in cluster A and two patients in cluster C had previously had tuberculosis.

Among the 12 incarcerated patients who were included in clusters, three episodes of transmission within the prison involving 6 patients, all of whom were HIV negative, were detected. The prison hospital confirmed a possible contact for each pair of patients by considering the incarceration period. Two clusters comprised two inmates, and the third cluster (cluster I; Table 1) included four patients, two of whom were in prison.

Homeless patients were included in 11 of the 26 clusters. According to the dates of diagnosis, these patients may have been the index cases in these 11 clusters, which involved a total of 50 patients. All of the homeless patients, both clustered and unclustered, presented with pulmonary tuberculosis.

**Risk factors associated with clustering.** The 95 clustered patients (group 2) were compared with the 177 unclustered patients (group 1) to identify the risk factors for inclusion in groups of tuberculosis transmission.

Table 2 presents the results of the univariate analysis. Patients in clusters were significantly more likely than patients with independent cases of infection to be male or homeless ($P < 0.0001$). Previous tuberculosis, pulmonary tuberculosis, and alcohol use were also associated with clustering ($P < 0.05$). Living in private homes was associated with a significantly lower risk of being in a cluster ($P < 0.001$). There was no significant association between clustering and birth in a foreign country, age, HIV status, incarceration, or residence in hostels for immigrant workers.

The associations with clustering identified in our univariate analysis and those identified in other reports were included in a multivariate analysis. Multivariate analysis showed that homelessness (odds ratio, 4.06; $P < 0.001$) and male sex (odds ratio, 3.21; $P = 0.003$) were significant independent risk factors associated with clustering. Eleven of 74 (14.9%) of the patients with neither of the two factors found to predict clustering in multivariate analysis (i.e., nonhomeless women) were clustered, whereas 43.3% (84 of 194) of patients with one factor found to predict clustering were clustered and 69.2% (27 of 39) of patients with both factors found to predict clustering belonged to clusters.

The association between homelessness and clustering varied according to the number of patients in the cluster (Fig. 1). Homeless patients were more frequently included in clusters involving a large number of patients. Nineteen of the 33 patients in the four largest clusters (clusters A to D) with 6 or more patients each were homeless. There were only 10 homeless people among the 62 patients included in clusters with less than six patients each. This difference is statistically significant ($P < 0.001$). No other risk factor associated with clustering was found for homeless patients.

## DISCUSSION

We performed a systematic, prospective, 1-year survey to define the risk factors associated with active tuberculosis transmission for the population living in Paris and surrounding area in 1995. It was based on the DNA fingerprinting of *M. tuber-*

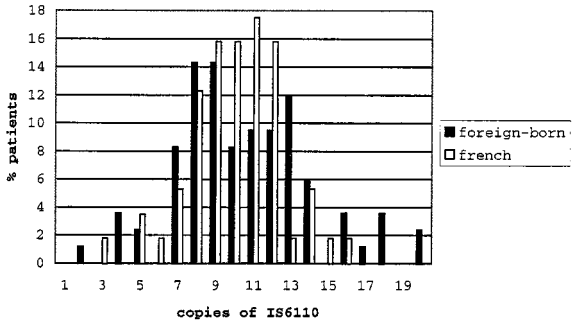


FIG. 2. Distribution of the number of IS*6110* copies in strains with unique RFLP patterns from 177 French and foreign-born patients with tuberculosis.

TABLE 2. Univariate analysis of risk factors for a patient to belong to a tuberculosis transmission group

| Characteristic | No. (%) of patients[a] | | Odds ratio | 95% Confidence interval | *P* value |
|---|---|---|---|---|---|
| | Group 1 | Group 2 | | | |
| Male sex | 108 (62.4) | 82 (86.3) | 3.8 | 1.87–7.83 | 0.00004 |
| Residence | | | | | |
| Homeless | 14 (10.4) | 29 (33.7) | 4.40 | 2.03–9.68 | 0.00002 |
| Hostel[b] | 12 (8.9) | 8 (9.3) | 1.05 | 0.37–2.94 | 0.9 |
| Prison | 17 (12.6) | 12 (14.0) | 1.13 | 0.47–2.68 | 0.77 |
| Private | 99 (68.2) | 37 (43.1) | 0.35 | 0.19–0.64 | 0.0002 |
| History of tuberculosis | 13 (8) | 17 (17.4) | 2.41 | 1.03–5.70 | 0.024 |
| Alcohol use | 16 (9.5) | 18 (19.1) | 2.25 | 1.02–4.98 | 0.026 |
| Pulmonary localization of TB[c] | 142 (82.1) | 88 (91.7) | 2.40 | 0.99–6.01 | 0.03 |
| Age, <60 yr | 134 (78.4) | 82 (85.4) | 1.62 | 0.78–3.38 | 0.16 |
| HIV positive | 22 (17.1) | 9 (14.1) | 0.80 | 0.31–1.99 | 0.6 |
| Foreign-born | 84 (59.6) | 46 (58.2) | 0.95 | 0.52–1.73 | 0.8 |

[a] Group 1, unclustered patients; Group 2, clustered patients.
[b] Hostel for immigrant workers.
[c] TB, tuberculosis.

*culosis* strains isolated from patients with tuberculosis diagnosed in several medical institutions in the area.

The combined use of different molecular markers gives a greater accuracy of strain clustering and more evidence of links between epidemiologically related patients. We used IS*6110* as a primary marker and the DR sequence as a secondary, independent marker for selected strains. The molecular markers revealed 74.3% genetic diversity. This rate is similar to that observed in other large cities in developed countries with populations of various geographical origins and where *M. tuberculosis* strains are imported from different geographical areas (1, 28).

In this study, we found that 35.7% of patients had isolates belonging to clusters, a proportion similar to that found in other studies in large cities in the United States (1, 15, 28). In a previous study conducted with isolates and patients from three of the hospitals participating in this study, there was evidence of transmission for only 13% patients (33). When we independently analyzed the results for these three hospitals, we found a proportion of clustered strains (14.3%; 11 of 76 patients) similar to that found in the previous study. However, when all the patients in the study were considered, we found that 19 of the 76 patients (25%) from these three hospitals were clustered. This demonstrates that a bias resulting in an underestimation of clustering can occur if the study is performed in a single institution and shows the importance of extending the study to a region, especially if a high level of extrahospital transmission is expected.

Confirmation of suspected cross contamination of specimens in mycobacteriology laboratories by molecular biology-based techniques has been reported previously (4, 27, 39). Our study reports that cultures for 24 (7.8%) patients were found to be falsely positive, and for 13 of these patients, the contamination was not suspected before strain fingerprinting. False-positive cultures can lead to an erroneous diagnosis of tuberculosis and may have serious consequences for the patients and their contacts (for example, inadequate treatment and the absence of a diagnosis of other diseases). Moreover, we detected two episodes of laboratory contamination involving falsely positive specimens from three patients and one patient, respectively, in which the source of contamination was a multidrug-resistant strain. Two of these patients had a previous history of tuberculosis with several relapses during the previous decade, and one patient had a chest X ray compatible with tuberculosis. False diagnosis of multidrug-resistant tuberculosis, even for patients who genuinely have tuberculosis, could lead to unnecessary treatment with second-line antibiotics. The occurrence of false-positive cultures is likely to increase as the population of patients at risk of *M. tuberculosis* infection increases. Typing of *M. tuberculosis* strains may avoid an erroneous diagnosis in both patients with a suspected and patients with an unexpected diagnosis. Strictly applied good laboratory practices and periodic fingerprinting as a quality control should be considered to avoid and detect cross contamination.

In this survey of tuberculosis patients, the statistically significant, independent risk factors contributing to patients' belonging to a transmission group were the type of residence and sex. By multivariate analysis, homelessness and male sex predicted clustering to a tuberculosis transmission group.

We expected to find associations between clustering and patients who rapidly progress from infection to active disease as a consequence of biological factors like HIV infection (5, 10, 29), as several reports have shown (1, 28), especially because the survey was planned for 1 year. In our study, HIV infection was not associated with clustering, even though 16% of the patients were HIV positive (similar to the average for tuberculosis patients in France). This may be related to the socioeconomic characteristics of the HIV-positive population in France, where 63% of the cases of HIV infection are attributed to sexual transmission and 26% are attributed to injection drug use (24). In our survey, clustered HIV-positive patients were significantly more likely to be included in small transmission groups with fewer than six patients. This shows that HIV-positive tuberculosis patients are less likely than non-HIV-infected tuberculosis patients to be sources of infection (15). This is partly because these patients are less contagious because they are less frequently smear positive and partly because the mortality rate among untreated HIV-infected patients with tuberculosis is higher (12). Some immunocompetent contacts may also have been infected by HIV-positive patients with no disease onset during the course of this 1-year survey.

In contrast to other reports (1, 15, 28, 41), we found that being a patient who was a native of the country under study was not associated with clustering. Birth in a foreign country was also not a risk factor for inclusion in a transmission group. Most foreign-born patients in this study were from African countries where the incidence of tuberculosis is high. We expected tuberculosis due to reactivation and a low rate of clustering for these patients. However, 38% of these patients were clustered. This reflects both their direct contacts with the French population and their difficult socioeconomic conditions, because 42% of the homeless patients in clusters were foreign-born.

There was no statistically significant difference for clustering between prison inmates and other patients. However, our data must be interpreted with caution. The rate of clustering depends on the size of the region covered by the hospitals, as described above. We detected three in-prison episodes of transmission, but the overall number of outbreaks across the region may be underestimated because only one prison hospital participated in the survey. The microbiology laboratory of this prison hospital performs the bacteriological diagnosis of tuberculosis for patients incarcerated in this prison and for

hospitalized patients from other prisons in Ile de France. Patients with tuberculosis from other prisons are treated locally and are not referred to the prison hospital if the patients do not need to be hospitalized. An evaluation of tuberculosis transmission in prison would require a targeted study with the participation of the medical staff of the different prisons of the region.

In our survey, the patient's age was not associated with independent or clustered cases, i.e., age was not more related to reactivation than to new infection. Tuberculosis in elderly people is usually considered to result from reactivation, and this has been demonstrated by several strain fingerprinting studies (1, 28, 41). However, about 30% of the patients older than 60 years had clustered isolates, showing that new infection may also be common in elderly people, as previously suggested by the reevaluation of primary resistance to antituberculosis drugs in France (20).

Epidemiological studies from several locations have shown that the incidence of tuberculosis in homeless people is 10 to 50 times that in the general population (21). Biological and social factors have been suggested as causes of this (13). Many contacts in this socioeconomic group are transient and are difficult to reconstruct by routine tracing techniques (28). These patients also poorly adhere to treatment, which is linked to longer periods for conversion to a negative culture, longer treatment regimens, and treatment failure (12, 23). Our study shows that more than 67% of tuberculosis cases in homeless patients were clustered, consistent with a study in Los Angeles, Calif. (3), and identified homelessness as a major risk factor for clustering. Transmission was found not only among homeless patients but also between homeless and nonhomeless patients, and homeless patients were probably the sources of infection for more than 50% of all patients in clusters. It shows that homeless patients play an important role in the transmission of tuberculosis. Tuberculosis control should focus on prevention and control of person-to-person transmission among homeless people and on the development of measures to make this population more aware of the threat of tuberculosis to improve their adherence to treatment. The success of such programs is a determinant to limiting the spread of *M. tuberculosis* throughout the community.

## ACKNOWLEDGMENTS

We are indebted to Pierre Nabeth (Epicentre) for assistance with statistical analysis.

This work was supported by the Association "Ensemble contre le SIDA." M. C. Gutiérrez was the recipient of a postdoctoral fellowship from the Spanish Ministry of Education and Culture.

## REFERENCES

1. **Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom.** 1994. Transmission of tuberculosis in New York City. N. Engl. J. Med. **24:**1710–1716.
2. **American Thoracic Society.** 1992. Control of tuberculosis in the United States. Am. Rev. Respir. Dis. **146:**1623–1633.
3. **Barnes, P. F., H. El-Hajj, S. Preston-Martin, M. D. Cave, B. E. Jones, M. Otaya, J. Pogoda, and K. D. Eisenach.** 1996. Transmission of tuberculosis among the urban homeless. JAMA **275:**305–307.
4. **Bauer, J., V. Thomsen, S. Poulsen, and A. B. Andersen.** 1997. False-positive results from cultures of *Mycobacterium tuberculosis* due to laboratory cross-contamination confirmed by restriction fragment length polymorphism. J. Clin. Microbiol. **35:**988–991.
5. **Bloch, A. B., I. M. Onorato, W. W. Ihle, J. L. Hadler, C. H. Hayden, and D. E. Snider.** 1996. The need for epidemic intelligence. Public Health Rep. **111:** 26–31.
6. **Bourdillon, F., B. Haury, and J. Salomon.** 1994. Situation de la tuberculose en Ile-de-France. Bull. Epidemiol. Hebd. **40:**185–187.
7. **Castel, O., C. Burucoa, B. Antoniotti, M. Underner, F. Clément, F. Patte, J. L. Fauchère, M. Castets, and V. Vincent.** 1994. Analyse d'une épidémie de tuberculose en 1992 dans le service de pneumo-phtisiologie du C.H.U. de Poitiers. Bull. Epidemiol. Hebd. **36:**165–167.
8. **Chaves, F., Z. Yang, H. E. Hajj, M. Alonso, W. J. Burman, K. D. Eisenach, F. Dronda, J. H. Bates, and M. D. Cave.** 1996. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **34:**1118–1123.
9. **Chevrel-Dellagi, D., A. Abderrahman, R. Haltiti, H. Koubaji, B. Gicquel, and K. Dellagi.** 1993. Large-scale DNA fingerprinting of *Mycobacterium tuberculosis* strains as a tool for epidemiological studies of tuberculosis. J. Clin. Microbiol. **31:**2446–2450.
10. **Daley, C. L., P. M. Small, G. F. Schecter, G. K. Schoolnik, R. A. McAdam, W. R. Jacobs, and P. C. Hopewell.** 1992. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus: an analysis using restriction-fragment-length-polymorphism. N. Engl. J. Med. **326:**231–235.
11. **Das, S., S. L. Chan, B. W. Allen, D. A. Mitchison, and D. B. Lowrie.** 1993. Application of DNA fingerprinting with IS*986* to sequential mycobacterial isolates obtained from pulmonary tuberculosis patients in Hong Kong before, during and after short course chemotherapy. Tubercle Lung Dis. **74:** 48–51.
12. **De Cock, K. M., N. J. Binkin, P. L. F. Zuber, J. W. Tappero, and K. G. Castro.** 1996. Research issues involving HIV-associated tuberculosis in resource-poor countries. JAMA **276:**1502–1507.
13. **Diez, E., J. Claveria, T. Serra, J. A. Cayla, J. M. Jansa, R. Pedro, and J. R. Villalbi.** 1996. Evaluation of a social health intervention among homeless tuberculosis patients. Tubercle Lung Dis. **77:**420–424.
14. **Drobniewski, F. A., R. J. Kent, N. G. Stoker, and A. H. C. Uttley.** 1994. Molecular biology in the diagnosis and epidemiology of tuberculosis. J. Hosp. Infect. **28:**249–263.
15. **Frieden, T. R., C. L. Woodley, J. T. Crawford, D. Lew, and S. M. Dooley.** 1996. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tubercle Lung Dis. **77:**407–413.
16. **Hermans, P. W. M., D. van Soolingen, J. W. Dale, A. R. J. Schuitema, R. A. McAdam, D. Catty, and J. D. A. van Embden.** 1990. Insertion element IS*986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J. Clin. Microbiol. **28:**2051–2058.
17. **Hermans, P. W. M., D. van Soolingen, E. M. Bik, P. E. W. de Haas, J. W. Dale, and J. D. A. van Embden.** 1991. The insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. Infect. Immun. **59:**2695–2705.
18. **Hermans, P. W. M., D. van Soolingen, and J. D. A. van Embden.** 1992. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. J. Bacteriol. **174:**4157–4165.
19. **Li, W. H.** 1981. Simple method for constructing phylogenetic trees from distance matrices. Proc. Natl. Acad. Sci. USA **78:**1085–1089.
20. **Marchal, G.** 1997. Recently transmitted tuberculosis is more frequent than reactivation of latent infections. Int. J. Tuberc. Lung Dis. **1:**192.
21. **Nolan, C. M., A. M. Elarth, H. Barr, A. M. Saeed, and D. Risser.** 1991. An outbreak of tuberculosis in a shelter for homeless men. A description of its evolution and control. Am. Rev. Respir. Dis. **143:**257–261.
22. **Otal, I., C. Martin, V. Vincent-Lévy-Frébault, D. Thierry, and B. Gicquel.** 1991. Restriction fragment length polymorphism analysis using IS*6110* as an epidemiological marker in tuberculosis. J. Clin. Microbiol. **29:**1252–1254.
23. **Pablos-Mendez, A., C. A. Knirsch, R. G. Barr, B. H. Lerner, and T. R. Frieden.** 1997. Nonadherence in tuberculosis treatment: predictors and consequences in New York City. Am. J. Med. **102:**164–170.
24. **Réseau National de Santé Publique.** 1997. Bulletin épidémiologique hebdomadaire, special edition. Epidémiologie des maladies à déclaration obligatoire en France: situation en 1995 et tendances évolutives récentes. Direction générale de la Santé, Paris, France.
25. **Ross, B. C., K. Raios, K. Jackson, and B. Dwyer.** 1992. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J. Clin. Microbiol. **30:**942–946.
26. **Sahadevan, R., S. Narayanan, C. N. Paramasivan, R. Prabhakar, and P. R. Narayanan.** 1995. Restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, India, by use of direct-repeat probe. J. Clin. Microbiol. **33:**3037–3039.
27. **Small, P. M., N. B. McClenny, S. P. Singh, G. K. Schoolnik, L. S. Tompkins, and P. A. Mickelsen.** 1993. Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J. Clin. Microbiol. **31:**1677–1682.
28. **Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik.** 1994. The epidemiology of tuberculosis in San Francisco. N. Engl. J. Med. **24:**1703–1709.
29. **Sudre, P., B. J. Hirschel, J. M. Gatell, S. Schwander, S. Vella, C. Katlama, B. Ledergerber, A. d'Arminio Monforte, F. D. Goebel, P. O. Pehrson, C. Pedersen, and J. D. Lundgren, and The AIDS in Europe Study Group.** 1996. Tuberculosis among European patients with the acquired immune deficiency

syndrome. Tubercle Lung Dis. **77:**322–328.
30. **Tchakamian, S., and B. Haury.** 1995. Les cas déclarés de tuberculose en 1994. Bull. Epidemiol. Hebd. **52:**225–227.
31. **Thierry, D., M. D. Cave, K. D. Eisenach, J. T. Crawford, J. H. Bates, B. Gicquel, and J. L. Guesdon.** 1990. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res. **18:**188.
32. **Torrea, G., G. Levée, P. Grimont, C. Martin, S. Chanteau, and B. Gicquel.** 1995. Chromosomal DNA fingerprinting analysis using the insertion sequence IS*6110* and the repetitive element DR as strain-specific markers for epidemiological study of tuberculosis in French Polynesia. J. Clin. Microbiol. **33:**1899–1904.
33. **Torrea, G., C. Offredo, M. Simonet, B. Gicquel, P. Berche, and C. Pierre-Audigier.** 1996. Evaluation of tuberculosis transmission in a community by 1 year of systematic typing of *Mycobacterium tuberculosis* clinical isolates. J. Clin. Microbiol. **34:**1043–1049.
34. **van Embden, J. D. A., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. W. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and P. M. Small.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.
35. **van Soolingen, D., P. W. Hermans, P. E. W. de Haas, D. R. Soll, and J. D. A. van Embden.** 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. **29:**2578–2586.
36. **van Soolingen, D., P. E. W. de Haas, P. W. M. Hermans, P. M. A. Groenen, and J. A. van Embden.** 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **31:**1987–1995.
37. **van Soolingen, D., L. Qian, P. E. W. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. A. van Embden.** 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. J. Clin. Microbiol. **33:**3234–3238.
38. **Wiid, I. J. F., C. Werely, N. Beyers, P. Donald, and P. D. van Helden.** 1994. Oligonucleotide $(GTG)_5$ as a marker for *Mycobacterium tuberculosis* strain identification. J. Clin. Microbiol. **32:**1318–1321.
39. **Wurtz, R., P. Demarais, W. Trainor, J. McAulney, F. Kocka, L. Mosher, and S. Dietrich.** 1996. Specimen contamination in mycobacteriology laboratory detected by pseudo-outbreak of multidrug-resistant tuberculosis: analysis by routine epidemiology and confirmation by molecular technique. J. Clin. Microbiol. **34:**1017–1019.
40. **Yang, Z. H., P. E. W. de Haas, D. van Soolingen, J. D. A. van Embden, A. B. Andersen.** 1994. Restriction fragment length polymorphism of *Mycobacterium tuberculosis* strains isolated from Greenland during 1992: evidence of tuberculosis transmission between Greenland and Denmark. J. Clin. Microbiol. **32:**3018–3025.
41. **Yang, Z. H., P. E. W. de Haas, C. H. Wachmann, D. van Soolingen, J. D. A. van Embden, and A. B. Andersen.** 1995. Molecular epidemiology of tuberculosis in Denmark in 1992. J. Clin. Microbiol. **33:**2077–2081.
42. **Yang, Z., I. Mtoni, M. Chonde, M. Mwasekaga, K. Fuursted, D. S. Askgard, J. Bennedsen, P. E. W. de Haas, D. van Soolingen, J. D. A. van Embden, and A. B. Andersen.** 1995. DNA fingerprinting and phenotyping of *Mycobacterium tuberculosis* isolates from human immunodeficiency virus (HIV)-seropositive and HIV-seronegative patients in Tanzania. J. Clin. Microbiol. **33:**1064–1069.
43. **Yuen, L. K. W., B. Ross, K. Jackson, and B. Dwyer.** 1993. Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. J. Clin. Microbiol. **31:**1615–1618.